

(19) 日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11) 特許番号

特許第3164346号  
(P3164346)

(45) 発行日 平成13年 5 月 8 日 (2001. 5. 8)

(24) 登録日 平成13年 3 月 2 日 (2001. 3. 2)

(51) Int.Cl.<sup>7</sup> 識別記号

A 6 3 H 3/33  
G 1 0 K 15/04  
G 1 0 L 15/00  
15/28

3 0 2

F I

A 6 3 H 3/33  
G 1 0 K 15/04  
G 1 0 L 3/00

C

3 0 2 F  
5 5 1 H  
5 7 1 H

請求項の数 8 (全 13 頁)

(21) 出願番号 特願平11-329009

(22) 出願日 平成11年11月19日 (1999. 11. 19)

(65) 公開番号 特開2000-325669 (P2000-325669A)

(43) 公開日 平成12年11月28日 (2000. 11. 28)

審査請求日 平成11年11月19日 (1999. 11. 19)

(31) 優先権主張番号 1 9 9 9 - 1 6 5 8 3

(32) 優先日 平成11年 5 月 10 日 (1999. 5. 10)

(33) 優先権主張国 韓国 (K R)

(73) 特許権者 399127359

株式会社韓国エキシス

大韓民国ソウル市瑞草区瑞草洞1624-2  
号

(72) 発明者 キム サンソウル

大韓民国ソウル市江南区開浦洞住公ア  
パート458棟302号

(72) 発明者 リュウ ジュウヒョン

大韓民国京畿道軍浦市山本洞乙支アパー  
ト611-401

(72) 発明者 カン ウォンイル

大韓民国京畿道始興市溪水洞493-12

(74) 代理人 100097515

弁理士 堀田 実 (外 2 名)

審査官 植野 孝郎

最終頁に続く

(54) 【発明の名称】 音声認識対話型人形おもちゃ

1

(57) 【特許請求の範囲】

【請求項 1】 人と動物の形態が混合した形状に形成された人形本体に、多数の文章のデジタル音声信号ストリームが所定の圧縮率で圧縮された音声圧縮データを記録している第1メモリ部 (33) と、外部から入力された使用者の音声信号を認識するための演算エリアが備えられている第2メモリ部 (35) とを備えた音声認識対話型人形おもちゃであって、

前記第2メモリ部 (35) に記録された、少なくとも1文章の使用者の音声を変換した音声信号に変換して出力し、伸張された音声信号を前記使用者に聴覚的に聞かせる音声入出力部 (37) と、

前記音声入出力部 (37) から出力されるフレーム単位の使用者のデジタル音声信号を一時的に記録するサーキュラバッファ (51) と、

2

前記サーキュラバッファ (51) に記録されたデジタル音声信号を、前記第1メモリ部 (33) に記録された圧縮データの音声認識用定数によって音声認識用単語に区分してビタビアルゴリズムで前記使用者の音声認識する音声認識部 (53) と、

前記音声認識部 (53) で認識された音声の内容が所定のシナリオに対応するように、少なくとも一つの応答文章を前記第1メモリ部 (33) で選択するダイアログマネージャ (55) と、

10 前記ダイアログマネージャ (55) で選択された前記第1メモリ部 (33) の音声圧縮データを伸張及び復元する音声デコーダ (57) と、

前記音声デコーダ (57) と音声入出力部 (37) 間に設けられた、アナログ音声信号とデジタル音声信号の一方を他方に変換する A/D・D/A コンバータ (47) と、

と、

前記第2メモリ部(35)とリストコントローラ(61)間に設けられ、前記第1メモリ部(33)のデータを前記第2メモリ部(35)に転送するメモリコントローラ(63)と、を含むことを特徴とする音声認識対話型人形おもちゃ。

【請求項2】 前記音声認識部(53)と前記第1メモリ部(33)との間と、前記ダイアログマネージャ(55)と前記第1メモリ部(33)との間には、前記第1メモリ部(33)から音声圧縮データと圧縮データの音声認識用定数を取り出し、第2メモリ部(35)に前記音声認識用データを転送するリストコントローラ(61)が備えられていることを特徴とする請求項1の音声認識対話型人形おもちゃ。

【請求項3】 前記音声認識部(53)は、前記サーキュラバッファ(51)に記録されたフレーム単位のデジタル音声信号から、前記第1メモリ部(33)の音声認識用定数によって所定の雑音を除去させ、一つの文字音声に対する固有値を特徴ベクトルとして算出する音声認識算出部(71)と、前記デジタル音声信号のサンプリング値から0点を検出するゼロクロッシングレート(73)と、前記ゼロクロッシングレート(73)での0点検出に対する信頼性を向上させるため、前記0点に対するエネルギーを算出するエネルギー算出部(75)と、前記ゼロクロッシングレート(73)と前記エネルギー算出部(75)の出力信号に基づいて、連続的なデジタル音声信号の中のいずれか1単語の端点データを検出する単位音声検出部(77)と、前記音声認識算出部(71)の特徴ベクトルデータと前記単位音声検出部(77)の端点データに基づいて1単語ずつ音声認識用単語に区分する前処理器(79)と、前記前処理器(79)で区分された単語に該当する第1メモリ部(33)の音声圧縮データが前記リストコントローラ(61)により取り出され、ビタビアルゴリズムで演算するようにした領域を提供する第2メモリ部(35)と、から成ることを特徴とする請求項1の音声認識対話型人形おもちゃ。

【請求項4】 前記人形本体の複数領域に埋設されており、使用者の接触を前記音声デコーダ(57)に知らせるため、人形本体の背中、鼻、口及び尻に設けられた接触スイッチ(T1、T2、T3、T4)を更に含む、ことを特徴とする請求項1の音声認識対話型人形おもちゃ。

【請求項5】 前記使用者が口、鼻、背中及び尻に設けられた接触スイッチ(T1、T2、T3、T4)と接触すると、それに対応する適切な音声を前記ダイアログマネージャ(55)と前記第1メモリ部(33)から取り出し、前記音声デコーダ(57)で実際の音声に伸張及び復元した後、前記音声入出力部(37)を通じて前記

4

使用者に聴覚的に聞かせる、ことを特徴とする請求項4の音声認識対話型人形おもちゃ。

【請求項6】 前記音声入出力部(37)は、前記使用者の音声と外部の雑音を電氣的信号に変換して前記サーキュラバッファ(51)に出力する第1マイクロフォン(39)と、前記外部の雑音を電氣的信号に変換して前記サーキュラバッファ(51)に出力する第2マイクロフォン(41)と、

10 前記音声デコーダ(57)で伸張及び復元された音声信号を電力増幅し、スピーカ(43)を通じて前記使用者に聴覚的に聞かせるための電力増幅部(45)と、から成ることを特徴とする請求項1の音声認識対話型人形おもちゃ。

【請求項7】 前記サーキュラバッファ(51)と前記第1及び第2マイクロフォン(39、41)との間に、該第1及び第2マイクロフォン(39、41)の出力信号をデジタルに変換し、前記音声デコーダ(57)と前記電力増幅部(45)間に、前記音声デコーダ(57)で伸張及び復元されたデジタル音声信号をアナログに変換するA/D・D/Aコンバータ(47)が設けられていることを特徴とする請求項6の音声認識対話型人形おもちゃ。

【請求項8】 前記A/D・D/Aコンバータ(47)と前記電力増幅部(45)との間には、使用者のボリュームを調節するための命令に応じて、前記電力増幅部(45)の出力強度を調節するボリューム制御部(49)が備えられている、ことを特徴とする請求項7の音声認識対話型人形おもちゃ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識対話型人形おもちゃに係り、より詳しくは、人形おもちゃ内に音声認識システムを設け、使用者と音声表現により面白い対話を行うことができる音声認識対話型人形おもちゃ及びその制御方法に関するものである。

【0002】

【従来の技術】一般に、子供は興味ある遊び又はおもちゃにより生活教育を学習する傾向があり、そのおもちゃとの親密な触れ合は、実社会へ導く模倣学習を実行するものである。このような模倣学習は、大部分は人形を通じて行われるものである。そこで、子供は自分で模倣学習のシナリオを作り、そのシナリオに従って、その人形に適切な反応を誘導し、即ち適切な音声表現と動作行為を双方向対話型に面白く進行することにより、その模倣学習に没頭することになる。

【0003】このように、おもちゃによる教育は、昔から子供に密着したものと引き継がれている。最近では、このような教育的効果を期待し得る発声人形の研究が活発になり、より進歩的な人形の製作が絶え間なく

試みられている。

【0004】このような従来の技術による人形おもちゃは、そのほとんどがタッチセンサが人形のある位置に設けられているものである。子供がこのタッチセンサを動作させると、磁気記録媒体（磁気テープ）又は半導体記録媒体（ICメモリ）に記録させた簡単な文章の音声表現、例えば「今日は、私はミッキーです。あなたは誰ですか。あなたは何をしていますか。」などのような不連続の短文の音声を発声するようになっていく。また、定型化された2、3通りの動作行為、例えば腕を上げる動作、頭を動かす動作などのような単純動作を行って、一時的な好奇心を満足させるに止めている。

【0005】

【発明が解決しようとする課題】従って、このような従来の人形おもちゃは、単発的で、簡単な文章を話す人形であり、タッチセンサの動作によって、シナリオのない単純な文章が録音された音声聞かせるため、一時的な好奇心を誘発することができる。しかし、直ぐに子供は飽きてしまい、実際にこのような人形おもちゃと遊ぶ期間が短くなるため、教育的な効果が低いという問題点がある。

【0006】また、従来の人形おもちゃが話す音声文章は、対話型のシナリオでなく、不連続的な文章の羅列であり、現実味に乏しいため、その教育的効果も次第に低下するという問題点がある。

【0007】本発明は、かかる問題点を解決するために創案されたものである。すなわち、本発明の目的は、子供等の使用者の音声認識し、子供の思考方式及び行動様式によって、少なくとも一つのシナリオに従って連続的に対話を行なえるようにした音声認識対話型人形おもちゃ及びその制御方法を提供することにある。

【0008】本発明の一番目の目的は、基本的に、話題に応じた音声出力を可能にし、子供がとり得る行動パターンをシナリオに作成し記録して、任意に設定された状況に応じて人形と双方向の対話を可能にすることにある。本発明の二番目の目的は、子供と対話する状況で、多様なシナリオに導くため、音声圧縮用ソフトウェアで音声を圧縮した後、これをロム（ROM）に記録し、必要時に速やかに取り出すことは勿論、一つの話題においても、選択可能な状況に応じて直ちに対話を可能にして、速やかな音声出力を可能にした装置に構成することにある。本発明の三番目の目的は、不特定多数人から入力される音声を理解するため、話者独立型という音声認識技法により、前記多数人の音声を学習させて、合理的な反応が起こるようにすることある。本発明の四番目の目的は、人形に触るときと撫でるときに備えた雑音研究により、適切に処理し得るソフトウェアで周辺の雑音と子供の音声とを識別し得るようにすることにある。本発明の五番目の目的は、4個の接触スイッチを備え、人形が特定姿勢となるか、子供が人形の所定部位に接触する

とき、即ち子供と人形との接触があるとき、適切な音声反応を通じて興味を抱かせることにある。本発明の六番目の目的は、入力された音声信号をシステムが理解し、これを適切に解釈して、実時間で適切な反応を行なえるようにハードウェアを具現化し、予め記録されたデータベースから、人が反応するような現実的な内容（シナリオ）を取り出して出力させ得るようにすることにある。

【0009】従って、本発明は、このような多様な機能と性能を充足させ得るように、先端のソフトウェアと先端の回路製作技術を実現するため、即ち人形に音声デコーダ、音声認識部、システムコントローラ、ダイアログマネージャー、その他の興味を誘発させ得る多様な補助機能を加えて、主に子供用人形おもちゃとしての外的及び内的要件を充足させ、言語による教育的効果（言語教育、遊び教育）を奏し得るよう、話者独立型、人工知能型、対話型の性能を有するようにした音声認識対話型人形おもちゃを提供することにその目的がある。

【0010】

【課題を解決するための手段】上記目的を達成するために、本発明によれば、人と動物の形態が混合した形状に形成された人形本体に、多数の文章のデジタル音声信号ストリームが所定の圧縮率で圧縮された音声圧縮データを記録している第1メモリ部（33）と、外部から入力された使用者の音声信号を認識するための演算エリアが備えられている第2メモリ部（35）とを備えた音声認識対話型人形おもちゃであって、前記第2メモリ部（35）に記録された、少なくとも1文章の使用者の音声を電氣的な音声信号に変換して出力し、伸張された音声信号を前記使用者に聴覚的に聞かせる音声入出力部（37）と、前記音声入出力部（37）から出力されるフレーム単位の使用者のデジタル音声信号を一時的に記録するサーキュラバッファ（51）と、前記サーキュラバッファ（51）に記録されたデジタル音声信号を、前記第1メモリ部（33）に記録された圧縮データの音声認識用定数によって音声認識用単語に区分してビタビアルゴリズムで前記使用者の音声を認識する音声認識部（53）と、前記音声認識部（53）で認識された音声の内容が所定のシナリオに対応するように、少なくとも一つの応答文章を前記第1メモリ部（33）で選択するダイアログマネージャ（55）と、前記ダイアログマネージャ（55）で選択された前記第1メモリ部（33）の音声圧縮データを伸張及び復元する音声デコーダ（57）と、前記音声デコーダ（57）と音声入出力部（37）間に設けられた、アナログ音声信号とデジタル音声信号の一方を他方に変換するA/D・D/Aコンバータ（47）と、前記第2メモリ部（35）とリストコントローラ（61）間に設けられ、前記第1メモリ部（33）のデータを前記第2メモリ部（35）に転送するメモリコントローラ（63）と、を含むことを特徴とする音声認識対話型人形おもちゃが提供される。

【0011】上記発明の構成では、話題に応じた音声出力を可能にし、子供が行う可能性のある行動パターンをシナリオに作成し記録させ、任意に設定された状況に応じて人形と双方向の対話が可能になる。例えば、子供と対話する状況で、多様なシナリオに導くため、音声圧縮ソフトウェアで音声を圧縮した後、これを第1メモリ部(33)に記録させ、必要時に速やかに取り出すことは勿論、一つの話題においても、選択可能な状況に応じて直ちに質疑応答が可能になる。また、この音声認識対話型人形おもちゃは、入力された音声信号をシステムが理解し、これを適切に解釈して、実時間で適切な反応を行なえるようにハードウェアを具現化し、予め記録されたデータベースから、人が反応するような現実的な内容(シナリオ)を取り出して出力させることができる。

【0012】

【0013】

【0014】

【0015】そして、前記音声認識部(53)と前記第1メモリ部(33)との間と、前記ダイアログマネージャ(55)と前記第1メモリ部(33)の間には、前記第1メモリ部(33)から音声圧縮データと圧縮データの音声認識用定数を取り出し、第2メモリ部(35)に前記音声認識用データを転送するリストコントローラ(61)が備えられていることが好ましい。

【0016】前記音声認識部(53)は、前記サーキュラバッファ(51)に記録されたフレーム単位のデジタル音声信号から、前記第1メモリ部(33)の音声認識用定数によって所定の雑音を除去させ、一つの文字音声に対する固有値を特徴ベクトルとして算出する音声認識算出部(71)と、前記デジタル音声信号のサンプリング値から0点を検出するゼロクロッシングレート(73)と、前記ゼロクロッシングレート(73)での0点検出に対する信頼性を向上させるため、前記0点に対するエネルギーを算出するエネルギー算出部(75)と、前記ゼロクロッシングレート(73)と前記エネルギー算出部(75)の出力信号に基づいて、連続的なデジタル音声信号の中のいずれか1単語の端点データを検出する単位音声検出部(77)と、前記音声認識算出部(71)の特徴ベクトルデータと前記単位音声検出部(77)の端点データに基づいて1単語ずつ音声認識用単語に区分する前処理部(79)と、前記前処理部(79)で区分された単語に該当する第1メモリ部(33)の音声圧縮データが前記リストコントローラ(61)により取り出され、ピタビアルゴリズムで演算するようにした領域を提供する第2メモリ部(35)と、から成るものである。

【0017】このような構成により、不特定多数人から入力される音声を理解するため、話者独立型という音声認識技法により、前記多数人の音声を学習させて、合理的な反応が起こるようにすることができる。

【0018】前記人形本体の複数領域に埋設されてお

り、使用者の接触を前記音声デコーダ(57)に知らせるため、人形本体の背中、鼻、口及び尻に設けられた接触スイッチ(T1, T2, T3, T4)を更に含むことが好ましい。

【0019】前記使用者が口、鼻、背中及び尻に設けられた接触スイッチ(T1, T2, T3, T4)と接触すると、それに対応する適切な音声を前記ダイアログマネージャ(55)と前記第1メモリ部(33)から取り出し、前記音声デコーダ(57)で実際の音声に伸張及び復元した後、前記音声入出力部(37)を通じて前記使用者に聴覚的に聞かせることが好ましい。

【0020】そして、前記音声入出力部(37)は、前記使用者の音声と外部の雑音を電氣的信号に変換して前記サーキュラバッファ(51)に出力する第1マイクロフォン(39)と、前記外部の雑音を電氣的信号に変換して前記サーキュラバッファ(51)に出力する第2マイクロフォン(41)と、前記音声デコーダ(57)で伸張及び復元された音声信号を電力増幅し、スピーカ(43)を通じて前記使用者に聴覚的に聞かせるための電力増幅部(45)と、から成るものである。

【0021】前記サーキュラバッファ(51)と前記第1及び第2マイクロフォン(39, 41)との間に、該第1及び第2マイクロフォン(39, 41)の出力信号をデジタルに変換し、前記音声デコーダ(57)と前記電力増幅部(45)間に、前記音声デコーダ(57)で伸張及び復元されたデジタル音声信号をアナログに変換するA/D・D/Aコンバータ(47)が設けられていることが好ましい。

【0022】このように、4個の接触スイッチ(T1, T2, T3, T4)を備え、人形が特定姿勢となるか、子供が人形の所定部位を接触するとき、即ち子供と人形との接触があるとき、適切な音声反応を通じて興味を抱かせることができる。

【0023】また、前記A/D・D/Aコンバータ(47)と前記電力増幅部(45)との間には、使用者のボリュームを調節するための命令(「音を大きく」及び「音を小さく」)に応じて、前記電力増幅部(45)の出力強さを調節するボリューム制御部(49)が備えられていることが好ましい。

【0024】

【0025】

【0026】

【発明の実施の形態】以下、添付図面に基づいて本発明の音声認識対話型人形おもちゃを詳細に説明する。図1は本発明の人形おもちゃを示す正面図、図2は本発明の人形おもちゃを示す側面図、図3は本発明の音声認識対話型人形おもちゃを示すシステムブロック図、図4は図3の処理順序による流れ図、図5は本発明のASIC化された音声認識部のブロック図、図6は本発明の音声認識対話型人形おもちゃの動作を示すフローチャートであ

る。

【0027】本発明の人形おもちゃは、図1及び図2に示すように、縫製人形の形状を有し、表皮が被せられており、頭部の内部と胴体の内部は、回路の保護のため、  
 ●堅固な枠（図示せず）を有する構造物からなっている。

【0028】図示例は、全体として人間に似ている妖精の形態を有しており、胴体は腹部及び胸部1と、4本の指を有する両手2、3と、2本の腕部8、9とから構成されている。下半身は2本の脚4、5と、4本の足指を有する両足6、7と、尻及び尾17から成る。顔は口10と、両耳11、12と、頭髮16と、両目14、15とから構成されている。

【0029】図2の側面図に示すように、頭部と胴体を連結する首19は屈曲可能な柔軟性材料から製作されており、頭部の回路と胴体との配線を容易にしてある。また、本発明の人形おもちゃは美しい外観を有すると共に、内部の回路を保護し得る素材で被覆してあり、柔らかく現実感を与える子供用人形の形状となるように形成されている。

【0030】そして、人形おもちゃには、これに接触した時に、人形の反応を誘導する接触スイッチが4箇所に配置してある。その配置位置は鼻（T1）、口（T2）、背中（T3）及び尻（T4）である。これらの接触スイッチT1、T2、T3、T4は、接触を容易に感知し得るように製作されたものである。これらの接触スイッチT1、T2、T3、T4は、非常に高い感度を有するものであり、縫製人形の皮膚内部に設けられ、接触したときに音声反応を誘導するため、制御部（ASIC；注文型半導体－マイクロプロセッサ）に直接ハイアクティブ信号が入力されるようになっている。特に、尻にある接触スイッチT4は、人形が立っているか座っているかを感知して、適合した反応を動作させる機能を有する。

【0031】例えば、尻の接触スイッチT4による反応は、うつ伏せにしている姿勢を取っているときは、「一緒に休みませんか」、立っている姿勢にあるときは、「遊びたい」、口の接触スイッチT2を触ると、「とてもおいしい」、この口の接触スイッチT2から手を離すと、「お腹がぺこぺこだ」、背中の接触スイッチT3を触ると、「誰ですか」、鼻の接触スイッチT1を触ると、「くすぐったい、アーアー」という音声出力される。

【0032】本発明の人形おもちゃのシステムは、図3に示すように、サーキュラバッファ51と、音声認識部53と、音声デコーダ57と、A/D・D/Aコンバータ47と、メモリコントローラ63を含む音声処理制御部（ASIC-application specific IC）30と、第1メモリ部（ROM）33と、第2メモリ部35と、音声入出力部37とを備えている。第1メモリ部33は、多数文章のデジタル音声信号ストリームが所定の圧縮率に圧

縮された音声圧縮データを記録しており、第2メモリ部35は、外部から入力された前記子供の音声信号を認識するための演算エリアを提供し、音声処理制御部30は、前記第2メモリ部35の演算エリアを用いて、前記子供の音声信号に相応する対話型応答を分析し、前記第1メモリ部33から、応答に相当する音声圧縮データを伸張及び復元し、音声入出力部37は、少なくとも1文章の前記子供の音声信号を電氣的音声信号に変換して前記音声処理制御部30に出力し、前記音声処理制御部30から、増幅された音声信号を前記子供に聴覚的に知らせる。

【0033】前記音声入出力部37は、図4に示すように、前記子供の音声と前記人形の表皮から発生する雑音を電氣的信号に変換して前記サーキュラバッファ51に出力する第1マイクロフォン39と、前記人形の表皮から発生する雑音を電氣的信号に変換して前記サーキュラバッファ51に出力する第2マイクロフォン41と、前記音声デコーダ57から、伸張及び復元された音声信号を電力増幅しスピーカ43を通じて前記子供に聴覚的に聞かせる電力増幅部45とを備えている。前記サーキュラバッファ51と前記第1及び第2マイクロフォン39、41との間には、前記第1及び第2マイクロフォン39、41の出力信号をデジタルに変換し、前記音声デコーダ57と前記電力増幅部45との間で、前記音声デコーダ57から、伸張及び復元されたデジタル音声信号をアナログに変換するA/D・D/Aコンバータ47が備えられている。ここで、前記スピーカ43は、前記第1メモリ部33に記録された圧縮音声信号が所定の過程により処理して使用者（子供）に聞かせるように構成されている。

【0034】一方、前記電力増幅部45の出力強度を調節して、実際に前記スピーカ43から発生する音声を大きくするためのボリューム制御部49は、前記A/D・D/Aコンバータ47と前記電力増幅部45との間に連結されている。例えば、前記ボリューム制御部49は、子供が所望の音声ボリュームに調節するため、前記第1マイクロフォン39を通じてボリューム調節のための命令（例えば、「音を大きく」及び「音を小さく」）が前記A/D・D/Aコンバータ47を介して入力されると、前記電力増幅部45を制御して、前記スピーカ43から、前記命令に応じたボリュームの音声が発生するように制御する。結果として、前記電力増幅部45は、前記音声処理制御部30のシステムコントローラ59のアンミュート（unMute）信号とボリューム制御部49の出力信号に基づいて、その大きさ及び利得が決定される。

【0035】また、前記音声入出力部37の第1及び第2マイクロフォン39、41は雑音除去機能を有するようにするためのもので、例えば、第1マイクロフォン39には音声と雑音が混ぜ合わされた信号が入力され、第2マイクロフォン41には、人形が使用者と接触する

か、周辺の雑音から影響を受けるとき、純粹雑音信号が入力される。この際に、音声処理制御部（ASIC）30は、両信号の雑音間相関関係（関数）を用いて雑音を減らすため、前記第1マイクロフォン39を通じた音声及び雑音信号と前記第2マイクロフォン41を通じた純粹雑音信号をコリレーション（CORRELATION）して雑音成分のみを除去させる。このような第1及び第2マイクロフォン39、41は、実験結果に基づいて、人形の両耳11、12に装着され、特に、前記第1及び第2マイクロフォン39、41のいずれか一つは小型のステレオマイクロフォンであり、そのステレオマイクロフォンとしては、音声周波数帯域において敏感で指向性の高いものが設けられている。

【0036】また、各接触スイッチT1、T2、T3、T4は、図4に示すように、音声処理制御部30に直接連結されている。前記サーキュラバッファ51は、前記音声入出力部37に入力された子供のデジタル音声信号、つまりA/D・D/Aコンバータ47で変換されたフレーム単位にデジタル化された音声サンプリング信号を一時的に記録し、音声認識部53は、前記サーキュラバッファ51に記録されたデジタル音声信号を、前記第1メモリ部33に記録された圧縮データの音声認識用定数によって、音声認識用単語に区分し、ヒタビアルゴリズムで前記子供の音声の意味を認識し、前記ダイアログマネージャ55は前記音声認識部53で認識された音声の内容が展開される多数のシナリオのいずれか一つを選\*

第1メモリ部の内部に記録された情報の内訳

記録内訳	種類	記録された量 (1word=16bit)	そのほか
圧縮された音	音声情報(160文章) 音楽(5曲) 子守唄(2曲) 話(5篇)	1,888kwords	約75分
音声デコード用データ	関数計算用定数	32kwords	15種
音声認識用データ	関数計算用定数	92kwords	9種

【0040】前記第2メモリ部35は、子供の音声及び応答文章の音声进行处理するための処理プログラムを記録して内部的なデータ信号処理を行う素子で、ブロックリスト（block list）のためのエリアと音声認識の前処理のための使用エリアを有し、所定のデータ記憶容量を有する。ここで、リストコントローラ61は、前記第2メモリ部35のデータを取り出す機能だけでなく、前記第1メモリ部33の圧縮音声データも取り出して音声デコード57に出力するようになっている。

【0041】ここで、前記第2メモリ部35と前記音声処理制御部30のリストコントローラ61との間には、相互間のデータ伝送のため、メモリコントローラ63が設けられている。特に、このメモリコントローラ63は前記圧縮データを読み取って前記第2メモリ部35に出力するように構成されている。

\*択し、その選択されたシナリオに対応するように、少なくとも1文章の圧縮音声データを前記第1メモリ部33から取り出し、前記音声デコード57は、前記ダイアログマネージャ55から取り出された前記音声圧縮データを伸張及び復元して前記音声出力部37に出力する。そして、前記第1メモリ部33、第2メモリ部35、ボリューム制御部49、A/D・D/Aコンバータ47、電力増幅部45に適切な制御信号を出力するシステムコントローラ59が設けられている。

【0037】また、前記音声処理制御部30は、前記子供が人形の鼻、口、背中、尻に設置された接触スイッチT1、T2、T3、T4に接触すると、それに対応する適宜の圧縮音声データを前記ダイアログマネージャ55及び前記第1メモリ部33から取り出し前記音声デコード57で変換して実際の音声に伸張及び復元した後、前記音声入出力部37のスピーカ43を通じて前記子供に聴覚的に聞かせるように制御する。

【0038】前記第1メモリ部33には、シナリオによる数多い文章の音声及び音楽、多数の話データ、音声認識用定数、音声デコーディングのための復元データが圧縮記録されている。使用された素子は4Mバイト以上の大容量を有し、1word単位（16ビット）にデータを記録して、総2Mwordsを記録することができる。記録された内容は次の表1に示す。

【0039】

【表1】第1メモリ部の内部に記録された情報の内訳

【0042】一方、電源供給部65は、3～24Vの電圧変動範囲にある任意の電圧を3.3Vの一定電圧に維持し、基本的に直列連結された乾電池3個（4.5V）の電圧を使用するが、その他の電源を使用することも可能である。このほかの動作のために必要な要素として、第2メモリ部35クロック用24.546MHzのクロック発生部67と、32.768kHzのタイマー69などの前記構成要素を動作させるのに必須のものであるので、その説明を省略する。

【0043】前記音声認識部53は、図5に示すように、前記サーキュラバッファ51に記録されたフレーム単位のデジタル音声信号から、前記第1メモリ部33の音声認識用定数によって所定の雑音を除去させ、1文字に対する固有値を特徴ベクトルデータとして算出する音声認識算出部71と、前記デジタル音声信号のサンプリ

ング値から0点を検出するゼロクロッシングレート73と、前記ゼロクロッシングレート73での0点検出に対する信頼性を向上させるため、前記0点に対するエネルギーを算出するエネルギー算出部75と、前記ゼロクロッシングレート73と前記エネルギー算出部75の出力信号に基づいて、連続的なデジタル音声信号のうち、どの1単語の端点データを検出する単位音声検出部77と、前記音声認識算出部71の特徴ベクトルデータと前記単位音声検出部77の端点データに基づいて、1単語ずつ音声認識用単語に区分する前処理器79と、前記前処理器79で区分された単語に該当する第1メモリ部33の音声圧縮データを前記リストコントローラ61により取り出しビタビアルゴリズムで演算する第2メモリ部35とを含んでいる。ここで、前記音声認識算出部71と前記サーキュラバッファ51との間には、前記サーキュラバッファ51のデジタル音声信号をより速やかに処理するために周波数増幅するブレンファシス81とが設けられている。

【0044】より詳しく説明すると、前記音声認識部53の計算の流れとモジュールの構成は、二つのモジュール群からなっており、ビタビアルゴリズムと単位音声検出アルゴリズムをASIC半導体で集積化させた多くのスーパーモジュールから構成されている。

【0045】まず、ビタビアルゴリズムは、4才から10才の年齢層の子供のおもちゃに使用し得るように、HMM (Hidden Markov Model) を用いるビタビアルゴリズムを使用して一つのチップから構成されている。また、ビタビアルゴリズムを実行する過程で生ずる多くの可変データを処理し得るよう、外部第2メモリ部35 (16Mバイト) で収容し得るブロックリスト構造を使用しており、全体的に約1Mバイトの第2メモリ部35の領域で動作するように構成されている。HMM学習方法は、使用者が異なっても、信頼性が向上されるようにし、即ち、話者独立型認識となるようにし、音素単位の認識をするものである。

【0046】図3、図4及び図5を参照して、前述した各要素の動作を簡略に説明すると次のようである。まず、二つのマイクロフォン39、41は音声信号を受けて電気信号に変換し、前述したA/D・D/Aコンバータ(Codec)47のアナログ音声信号変換装置に送る。この際に、入力された二つの音声信号は、雑音除去のため、互いに独立した形態で音声処理制御部30に伝達されて、コリレーションを行う。前記音声処理制御部30では、特別な状況がない限り、A/D・D/Aコンバータ47に制御信号(データ入力準備信号)を送出して、収容準備状態になったことを知らせ、A/D・D/Aコンバータ47で補間(Interpolation)のためにx256FSの値である2,048MHzを使用し、その同期周波数(SYNC Frequency)は8kHzであり、音声認識部53で音声の認識を向上させるサンプリング率で適用され

るようにする。特に、前記8kHzサンプリング率は、前記音声処理制御部30の音声認識部53において、認識アルゴリズムに対する重要な処理基準となっている。一方、入力された音声信号はA/D・D/Aコンバータ47でA/D変換されて音声処理制御部30に送られ、第1及び第2マイクロフォン39、41を通じて独立データとして入力されて、前記コリレーション演算によるノイズがフィルタリングされる。

【0047】このように、雑音の除去されたデジタル音声サンプリング信号は、サーキュラバッファ51でフレーム単位で一時的に記録されてから、ブレンファシス81と音声認識算出部71から一つ一つの使用者の音声に対する固有値が特徴ベクトルとして算出され、各々の単語の端点を検出するため、ゼロクロッシングレート73及びエネルギー算出部75と単位音声検出部77をほぼ同時に経ることになり、これらのそれぞれは前処理器79で単語ずつ音声認識用単語に区分される。すると、リストコントローラ61から前記前処理器79からの音声認識用単語に該当する第1メモリ部33の圧縮データを取り出すと、第2メモリ部35にこれらのデータとビタビアルゴリズムを移して、認識のための演算動作を行って分析する。

【0048】より詳しく説明すると、実際に、その動作は8kbpsでサンプリングされた音声信号→前処理(音声特徴検出)→音声検出→音声認識の段階からなっている。前処理は、Power、Hamming Window、ブレンファシスなどの計算段階を経た後、リアルエフエフティー(RealFFT)をしたスペクトル結果に対してメル(Mel)スケールのケプストラム(Cepstrum)を計算する。これとは別に、音声をゼロクロッシングレート73とエネルギー算出部75で計算して、音声の始点と端点を検出する。このような二つの音声検出結果に基づいて、音声認識の開始及び終了あるいはリセットの有無を決定し、メルスケールケプストラム係数列とHMMに対してビタビアルゴリズムを用いて遂に音声を認識する。勿論、このような数多い計算をするために必要な定数は第1メモリ部33に記録されてから、必要となるたびに取り出されて使用される。また、必要な値を計算してから取り出す作業のため、第2メモリ部35を使用し、そのデータ計算の膨大性のため、リストコントローラ61を用いている。ここで、音声認識及び圧縮のため、音声の端点検出は認識率と圧縮率を高めるのに使用される単位音声検出部77でなされる。

【0049】ところで、ゼロクロッシングレート73とエネルギー算出部75は実験室又は比較的静かな室内で高い効率的中率を表すが、僅かな騒音にも反応するスピーチ端点検出では根本的な問題点を有していることが事実であるので、メルケプストラムと共に動作されなければならない。

【0050】即ち、音声、雑音、無音の混ぜ合わせられ



たサンプリング信号をエネルギー検出部、ゼロクロッシングレート、メルスケールケプストラムを求めて単位音声検出部に入力すると、音声（雑音混合）部分が出力される。このように、二つのモジュールから出た結果は前処理器79に送られて、音声信号を認識することになる。

【0051】このように、音声認識部53で使用者、つまり子供の音声認識されると、ダイアログマネージャ55は、その認識された音声を多数のパターンに分けられたシナリオのいずれか一つを選択し、その選択された一つのシナリオによる応答音声の圧縮データを前記リストコントローラ61と前記第1メモリ部33から取り出して音声デコーダ57に伝達する。

【0052】次いで、音声デコーダ57は、前記第1メモリ部33の圧縮データを所定のデコーディング過程により伸張させてデジタル音声信号に復元し、音声入出力部37を通じて話者である子供に聞かせる。この際に、前記音声デコーダ57と前記音声入出力部37の間にはA/D・D/Aコンバータ47が備えられているため、デコーディングされた前記デジタル音声信号はアナログに変化されて実際音声として発生される。

【0053】ここで、ボリュームを調節するため、音声入出力部37を通じて、子供の音声から、「音を大きくしろ」という命令が入力される場合、この命令はA/D・D/Aコンバータ47を介して音声処理制御部30に入力されて認識される。すると、音声デコーダ57のボリューム制御信号に応じて、所定の利得が決定されたボリューム制御部49は前記A/D・D/Aコンバータ47に出力される前記アナログ音声信号を通常の増幅利得値より大きくして子供の耳に聞かせるように、前記電力増幅部45を制御する。

【0054】前述したような本発明による音声認識対話型人形おもちゃは、図6に示すような制御方法により操作する。まず、人形が反応し得る段階は3段階（バッテリーオン、タイムシグナル、接触スイッチモード）に区分される（130、131、132）。仮に、電源が供給されるか、電源の電池が取り替えられると、予め決められた挨拶のことは、つまり「こんにちは、私はサラです。あなたは誰ですか。」という挨拶の言葉が出力される（130）、使用者が録音しておいたメッセージ、つまり朝、昼、夕によってそれぞれ相違したメッセージの挨拶の言葉、又は設定されたシナリオによる言葉が音声として出力され（タイマーモード、メッセージモード）（131）、接触スイッチT1、T2、T3、T4により、挨拶の言葉に対応する音声出力されることもある（132）。

【0055】この流れ図によると、初期化により挨拶の言葉出力した後（133）、子供の対話用音声を認識するため、音声信号を待つ（134）。仮に、子供が応答しない待機時間が長くなると（137）、時間を終了し（144）、まず、待機中であることを知らせる擬声

語、歌モード、話モード、遊びモード（病院ごっこ、ままごと、市場遊び、ティーパーティ遊び）の中の任意の状況を付与するか、又はこれらのモードに対する案内音声出力する（145）。この際に、待機時間は約10秒程度である。このような過程が最小限3回以上であると、セーブモード状態となるが（147）、この状態にないときは、続けて子供の音声を待つ（146）。

【0056】次いで、子供が望む音声反応をすると、すぐ、挨拶の言葉を出力し、上述した各遊びに関する所望の遊びモード（136、138）に動く（141）。仮に、応答がないか、認識が不可能（未認識）である場合は、決められた方式の質問が繰り返される（143）。所望の遊びモードが認識されたときは、すぐ、開始を知らせる音声出力され（148）、遊びが開始される。この段階を経た後、多様なパターンによって、使用者と人形は続けて進行することができる（149、150、151、153、154）。仮に、対話中に認識が不可能であっても、人が相手の言葉を認識し得ない場合に類似した行動を見せるように、決定アルゴリズムを作って計算した後、再び問い合わせるか、意図的にパターン上で可能な応答をするか、再び遊びを進行させるかを決定する（153、157）。これは人形に内蔵した処理方式に依存する。多様なパターンを経て遊びが終わると、使用者は再びほかの遊びをするか、止めるかを選択すべきである（155、156、158）。このような遊び方式が4種存在し、可能なパターンの数は約3,000種に至る。

【0057】使用者にもっと興味を与え、実用性を強調するため、タイマーモードとメッセージ録音及び再生モードがある。すべての調整は、人形の内部に記録された音声のみによってなされ、使用者はセッティングの正確性を出力される音声により判断し得るようになってい

【0058】タイマーモードは四つの接触スイッチによる3ビットの信号を用い、七つのセッティングモード、八つのモード、時刻調整モードがある。また、電力の供給が中断されて、使用者がセッティングしたデータが消えるときに備えて、デフォルトセッティング機能がある。セッティングモードは、一度押すとセッティング状態がオンとなり、二度押すとオフとなる。また、5秒以上何の操作もしないと、自動的に元の状態を記録したままでもオフされる。セッティング可能なモードとしてはノーアクションモード、時間調整モード、起床時間、朝食時間、昼食時間、昼寝時間、夕食時間、就寝時間を知らせる八つのモードがあり、各モードで、使用者の音声メッセージが決められた時刻に出力されるようにセッティングすることができる。この際に、時刻調整モードは、分単位にセッティング可能であり、修正された後、音声で結果を知らせる。もちろん、このような機能の一部のみを動作させたいときは、セッティングモードを順次押



して、オン／オフさせることができる。仮に、中途に実行を止めても、自動的に既存に実行された値を記録し、オフされる。モード調整スイッチを一度押すと、時刻修正を知らせる音声が出力され、もう一度押すと、起床時間をセッティングするモードとなり、残りも同一形式でなされる。

【0059】メッセージ録音（再生）モードを用いて使用者の音声を4分程度録音することができ、それ以上の録音をすると、自動的に中断される。一方、八つのモードは所望の音声を七つの使用者セッティング時間に合わせ10秒間に調整が行われないと、自動的にオフされる。仮に、中間に実行を止めても、自動的に実行された値を記録しオフされる。

【0060】圧縮された音声データを音声信号に変換する音声デコーダ57は第1メモリ部33に記録された圧縮音声情報（144ビット／240サンプル 16kHz サンプル データ）を与えられたアルゴリズムでデコーディングするモジュールで、総14個のサブモジュールから構成され、次の順に従って進行される。第2メモリ部35に初期値を付与し（set\_init）、第1メモリ部33の圧縮データを読み取り（rd\_dat）、monotone、lsf\_intを用いてLSF処理をした後、con\_gain、stoch\_cw、adapl\_cw、adap2.cw、lsf\_pc、lp\_syn、post\_fitの処理モジュールを経て、音声信号が出力される。

#### 【0061】

【発明の効果】上述したように、本発明の音声認識対話型人形おもちゃは、音声認識手段と音声発声手段から構成されたシステムと、シナリオ展開を可能にするダイアログマネージャとを結合させて対話用に構成したものであり、子供のための人形であることを考慮して機械構造のシステムを縫製した人形内に入れて、この人形おもちゃと遊びたいという欲求を誘発し、かつ言語の教育的効果を高める、等の優れた効果がある。

#### 【図面の簡単な説明】

【図1】本発明の人形おもちゃを示す正面図である。

【図2】本発明の人形おもちゃを示す側面図である。

【図3】本発明の音声認識対話型人形おもちゃを示すシステムブロック図である。

【図4】図3の処理順序による流れ図である。

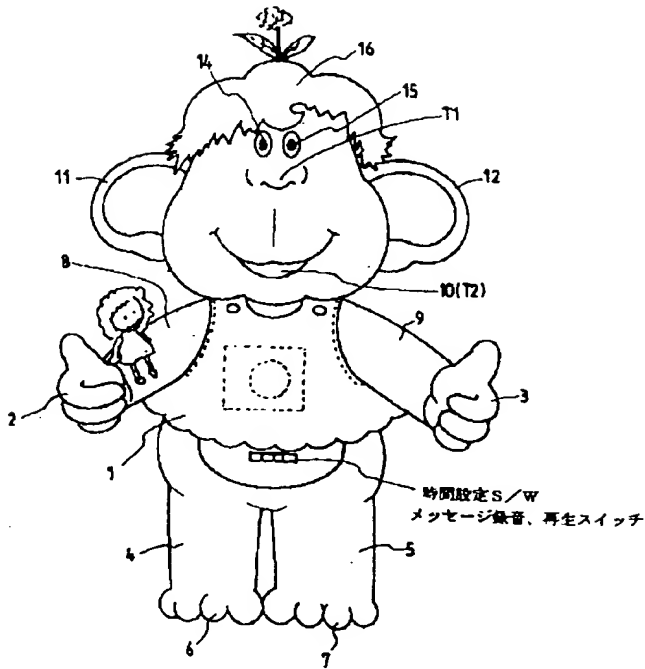
【図5】本発明のASIC化された音声認識部のブロック図である。

【図6】本発明の音声認識対話型人形おもちゃの動作を示すフローチャート図である。

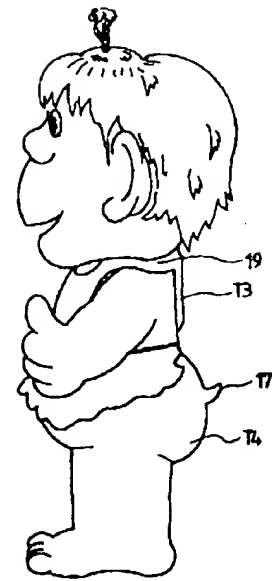
#### 【符号の説明】

- 30 音声処理制御部
- 33 第1メモリ部
- 35 第2メモリ部
- 37 音声入出力部
- 39 第1マイクロフォン
- 41 第2マイクロフォン
- 43 スピーカ
- 45 電力増幅部
- 47 A/D・D/Aコンバータ
- 49 ボリューム制御部
- 51 サーキュラバッファ
- 53 音声認識部
- 55 ダイアログマネージャ
- 57 音声デコーダ
- 59 システムコントローラ
- 61 リストコントローラ
- 63 メモリコントローラ
- 65 電源供給部
- 67 クロック発生部
- 69 タイマー
- 71 音声認識算出部
- 73 ゼロクロシングレート
- 75 エネルギー算出部
- 77 単位音声検出部
- 79 前処理器
- 81 プレエンファシス
- T1, T2, T3, T4 接触スイッチ

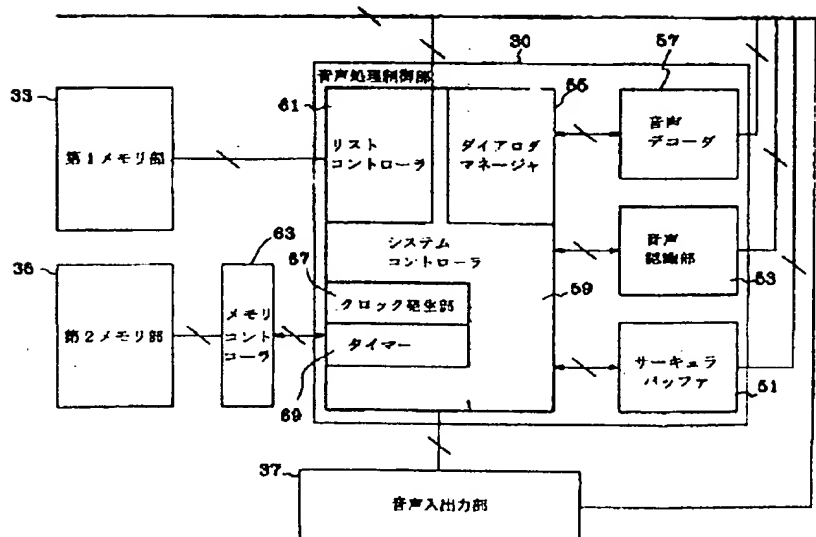
【図 1】



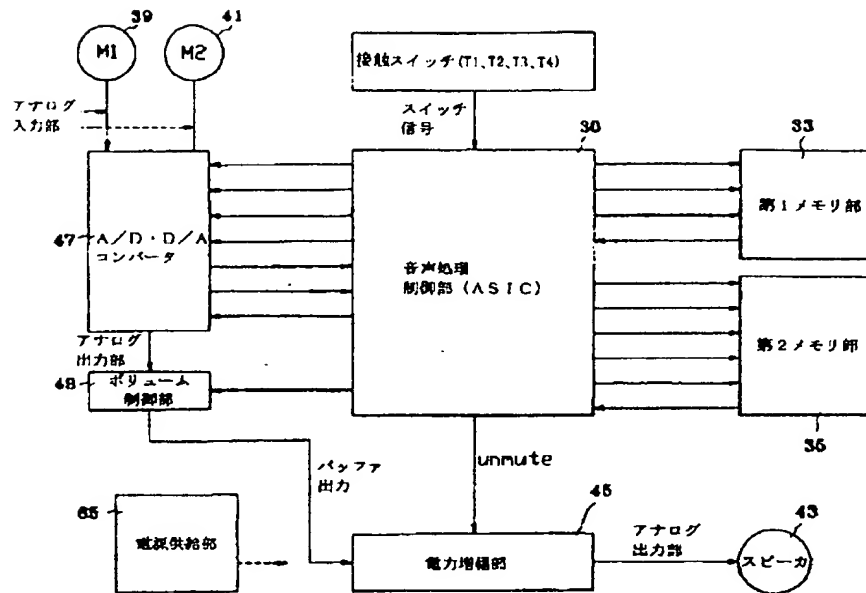
【図 2】



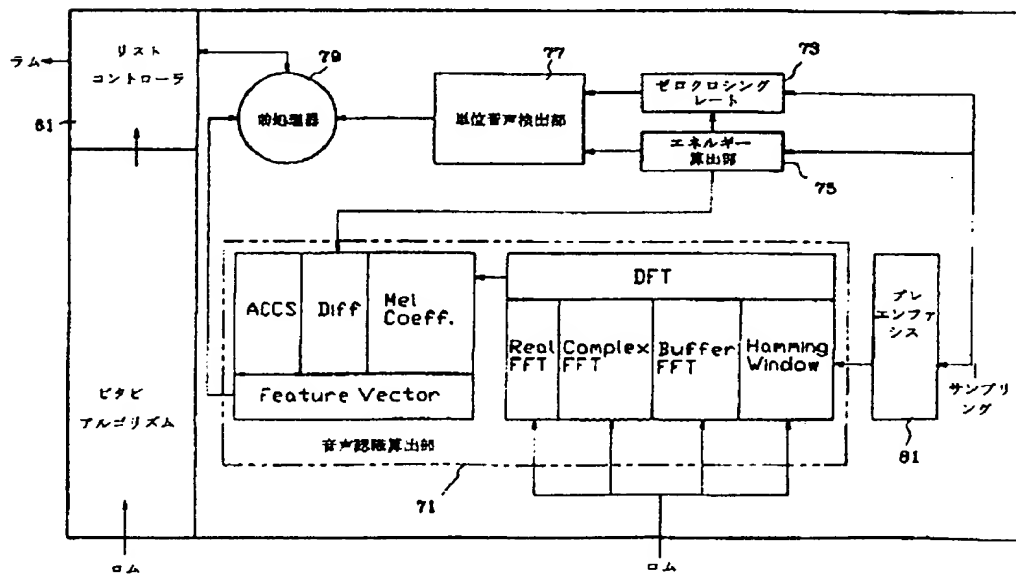
【図 3】



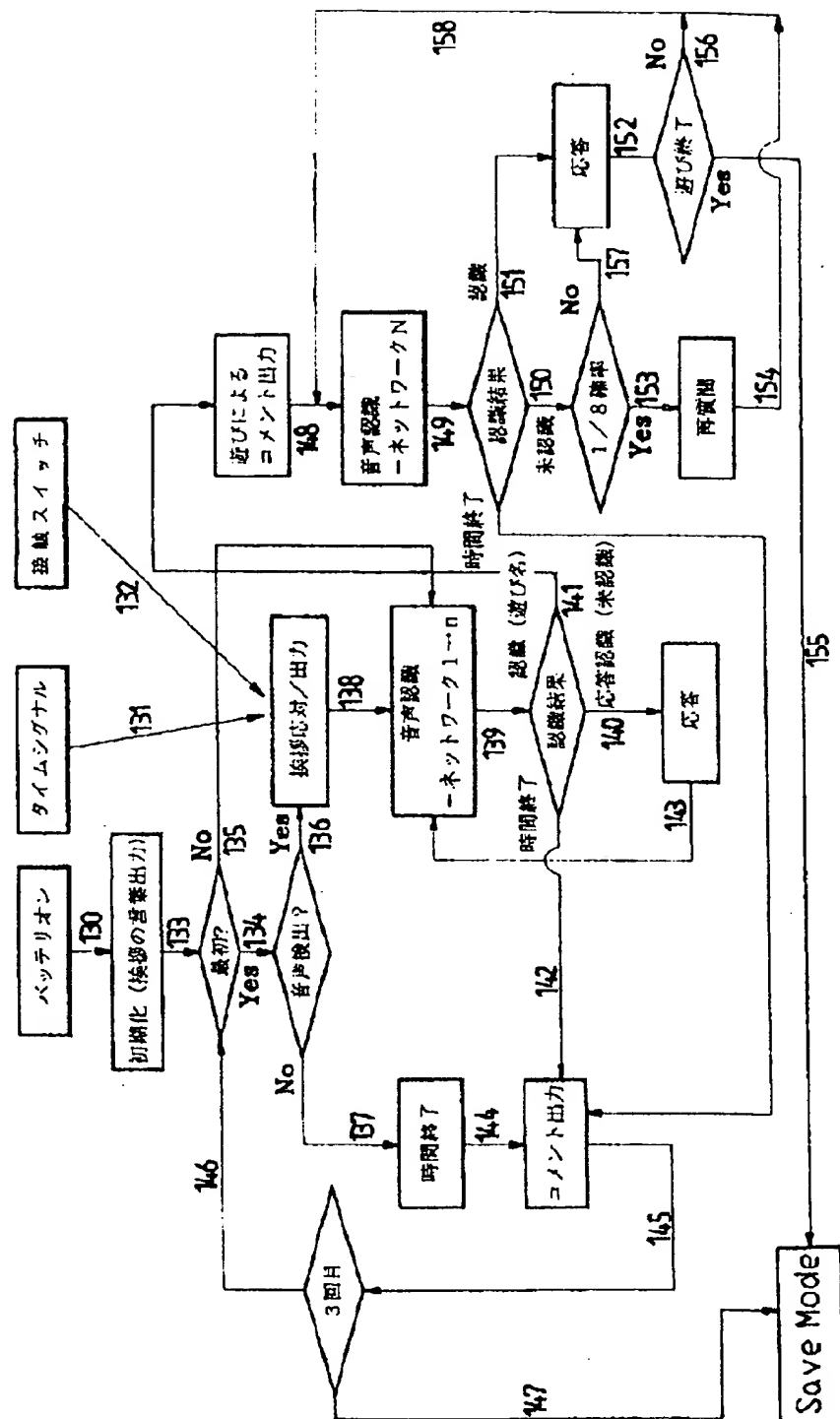
【図4】



【図5】



【図 6】



## フロントページの続き

(72)発明者     パク   ヨンジョン  
                  大韓民国京畿道波洲市内面白連里28

(72)発明者     キム   ウンジャ  
                  大韓民国ソウル市馬浦区桃花洞現代2次  
                  アパート208-1201

(72)発明者     クオン   サッボン  
                  大韓民国ソウル市江西区藤村洞公アパー  
                  ト805-402

(72)発明者     イー   チェキョン  
                  大韓民国全羅南道光州広域市西区花正1  
                  洞786-22

(72)発明者     チー   キョンチェ  
                  大韓民国慶尚南道馬山市廻円区廻円2洞  
                  509-11

(72)発明者     パン   テイシク  
                  大韓民国京畿道南陽州市錦谷洞310-1

(72)発明者     ハン   チュイヨン  
                  大韓民国京畿道高陽市徳陽区幸新1洞ヘ  
                  ッピッマウル1904棟902号

(56)参考文献     特開   平10-179941 (J P, A)  
                  特開   平9-326856 (J P, A)

(58)調査した分野(Int.Cl.<sup>7</sup>, D B名)

A63H	1/00 - 37/00
G10K	15/04
G10L	15/00
G10L	15/28